



شرکت آب منطقه‌ای مازندران
کمیته تحقیقات و پژوهش‌های کاربردی

خلاصه گزارش طرح
منطقه‌ای کردن حوضه آبریز استان مازندران به منظور پیش‌بینی سیلاب

کد طرح
MAE-91011

پژوهشگر
دکتر علیرضا مقدم‌نیا (پردیس کشاورزی و منابع طبیعی دانشگاه تهران)

ناظر طرح
دکتر وحید غلامی و گروه آب‌سطحی و قنات‌مطالعات پایه

تاریخ انتشار
تابستان ۱۳۹۸

شناسنامه طرح

این گزارش نتیجه اجرای طرح تحقیقاتی با عنوان «**منطقه‌ای کردن حوضه آبریز استان مازندران به منظور پیش‌بینی سیلاب**» می‌باشد که پیشنهاد انجام آن پس از داوری در کمیته تحقیقات آب منطقه‌ای مازندران مورد تایید قرار گرفت.

قرارداد این طرح در تاریخ ۹۲/۰۴/۲۹ به شماره ۱۰۹۰۹/۵۰۰ بین شرکت آب منطقه‌ای مازندران به عنوان کارفرما و دانشگاه تهران به عنوان مجری (مشاور) تنظیم و با حمایت و پشتیبانی مالی از محل اعتبارات جاری شرکت آب منطقه‌ای مازندران انجام شده و در تاریخ ۹۸/۵/۳۱ خاتمه یافته است.

پژوهشگر این طرح آقای دکتر علیرضا مقدم‌نیا بوده، داوری نهایی طرح توسط آقای دکتر وحید غلامی و گروه آب سطحی دفتر مطالعات پایه انجام گردیده است.

کمیته تحقیقات و پژوهش‌های کاربردی
شرکت آب منطقه‌ای مازندران

منطقه‌ای کردن حوزه آبریز استان مازندران به منظور پیش‌بینی سیلاب

پژوهشگران: علیرضا مقدم‌نیا، فرهاد فرسادنیا و محسن رستمی کامرود

کد طرح: MAE-91011

چکیده

یکی از روش‌های تخمین چندک‌های سیلاب در حوزه‌های فاقد آمار یا با طول دوره آماری کوتاه، استفاده از روش تحلیل فراوانی منطقه‌ای است. در مطالعات تحلیل فراوانی منطقه‌ای، به‌منظور دستیابی به مناطق همگن هیدرولوژیک از تکنیک‌های خوشه‌بندی استفاده می‌شود. اخیراً در چند تحقیق از نگاشت ویژگی خود سامان (Self-Organizing Feature Maps) استفاده شده است. اما مشکل اصلی SOM تفسیر نقشه خروجی از این روش به‌منظور یافتن مناطق همگن هیدرولوژیک است. به این دلیل از SOM به‌عنوان ورودی الگوریتم‌های خوشه‌بندی دیگر به‌کار می‌رود. در این مطالعه، ابتدا از نگاشت ویژگی خود سامان برای شکل‌گیری یک نقشه ویژگی دو بعدی استفاده شد، سپس گره‌های خروجی از نگاشت ویژگی خود سامان توسط الگوریتم خوشه‌بندی C- میانگین فازی برای شکل‌گیری مناطق مورد نیاز در تحلیل فراوانی منطقه‌ای سیلاب استفاده شد. تعداد بهینه خوشه‌های فازی براساس شش شاخص ارزیابی خوشه‌بندی فازی تعیین شد. نتایج نشان داد که شاخص‌های زی-بنی تعمیم یافته و وون نسبت به شاخص‌های دیگر کارایی بهتری دارند. سپس همگنی هر یک از مناطق بدست آمده از الگوریتم خوشه‌بندی ترکیبی توسط آماره ناهمگنی هاسکینگ و والیس ارزیابی شد. همچنین آماره Z^{DIST} برای ایستگاه‌های هر یک از چهار منطقه‌ای همگن هیدرولوژیک، برای توزیع‌های ۳ پارامتری محاسبه شد و توزیع لجستیک تعمیم یافته به‌عنوان بهترین توزیع منطقه‌ای انتخاب شد. پارامترهای این توزیع به‌وسیله روش گشتاورهای خطی محاسبه شد، سپس سیلاب با دوره بازگشت‌های مختلف در هریک از ایستگاه‌ها با استفاده از روابط منطقه‌ای برآورد شد. با توجه به نتایج این پژوهش، می‌توان از الگوریتم ارایه شده برای دستیابی به مناطق همگن هیدرولوژیک در نقاط دیگر کشور استفاده نمود. در ضمن یک برنامه رایانه‌ای کاربرپسند برای خوشه‌بندی SOM و فازی و نیز محاسبه شاخص‌های ارزیابی خوشه‌بندی فازی به‌صورت گرافیکی توسط پژوهشگران این تحقیق در محیط ویندوز توسعه یافت.

کلید واژه‌ها: منطقه‌ای کردن، تحلیل خوشه‌ای فازی، نگاشت ویژگی خود سامان، سیلاب شاخص، گشتاورهای خطی، همگنی هیدرولوژیک.

مقدمه

در هیدرولوژی برآورد فراوانی و بزرگی مقادیر حد طبیعی مانند سیلاب، جریان کمینه و حداکثر بارندگی اهمیت فراوانی دارد. به دلیل نادر بودن مقادیر حدی و کوتاه بودن طول دوره آماری، تخمین فراوانی وقوع این حوادث دشوار است، بنابراین از تحلیل فراوانی منطقه‌ای برای تخمین واقع بینانه‌ی چندک‌ها^۱ استفاده می‌شود. در روش تحلیل فراوانی منطقه‌ای، باید ایستگاه‌ها به مناطق همگن اختصاص یابند، زیرا همگنی منطقه برای اطمینان از قابل اعتماد بودن تحلیل منطقه‌ای الزامی است (Hosking and Wallis, 1997). زمانی که در منطقه تعداد ایستگاه‌های زیادی برای تحلیل فراوانی منطقه‌ای وجود نداشته باشد، شناسایی مناطق همگن هیدرولوژیکی دشوار می‌باشد. به این دلیل هیدرولوژیست‌ها از تحلیل خوشه‌ای برای دستیابی به مناطق همگن هیدرولوژیک استفاده می‌کنند. روش‌های خوشه‌بندی متعددی برای طبقه‌بندی حوضه‌های آبخیز ارائه شده است. از جمله این روش‌ها می‌توان به خوشه‌بندی میانگین k (Burn and Goel, 2000) و خوشه‌بندی سلسله مراتبی تراکمی (Hosking and Wallis, 1997) و خوشه‌بندی هیبرید (Rao and Srinivas, 2006) اشاره نمود. الگوریتم نگاشت ویژگی خود سامان (Kohonen, 1982) یک مدل اکتشافی برای مجسم ساختن و کشف روابط خطی و غیر خطی در مجموعه داده‌ها با ابعاد زیاد است. SOM کاربرد زیادی در کارهای صنعتی مانند شناسایی گفتار، مدل‌سازی زیستی، فشردن داده‌ها، پردازش سیگنال و داده کاوی دارد (Kohonen, 2001). اخیراً نگاشت ویژگی خود سامان به عنوان یک ابزار انفورماتیک مدرن در شناسایی مناطق همگن هیدرولوژیک به کار رفته است. (Hall and Minns (1999) از SOM برای منطقه‌ای کردن ایستگاه‌های آب‌سنجی در جنوب غربی انگلستان و ولز با استفاده از مساحت حوضه آبخیز، طول آبراهه اصلی، شیب آبراهه اصلی، میانگین بارندگی سالانه و شاخص خاک استفاده کردند. آنها نرون‌های خروجی از SOM را به سه گروه مشخص به منظور دستیابی به سه منطقه‌ی همگن تقسیم کردند. (Lin and Wang (2006) یک روش تک مرحله‌ای را برای انجام تحلیل خوشه‌ای براساس SOM ارائه کردند. آنها این روش را برای فاکتورهای هیدرولوژیکی تاثیر گذار در منحنی‌های مدت جریان کمینه در جنوب تایوان به کار بردند. (Jingyi and Hall (2004) روش‌های خوشه‌بندی وارد، خوشه‌بندی فازی و شبکه عصبی کوهونن را برای طبقه‌بندی حوضه‌ی رودخانه گان-مینگ در جنوب شرق چین بکار بردند. نتایج آنها نشان داد که شبکه عصبی کوهونن نسبت به دو روش دیگر از

1 - Quantile

توانایی‌های بیشتری در تعیین زیر منطقه‌های همگن هیدرولوژیک برخوردار است. (Lin and Chen (2006 از الگوریتم SOM برای خوشه‌بندی ۱۵۴ ایستگاه باران‌سنجی در تایوان به منظور تحلیل فراوانی منطقه‌ای استفاده کردند. آنها از ۱۷ خصوصیت حوضه‌های آبخیز برای خوشه‌بندی استفاده کردند و اندازه نقشه خروجی را به منظور مطمئن شدن از حداکثر تعداد خوشه، 12×12 انتخاب کردند و در نهایت منطقه مورد بررسی را به ۸ خوشه همگن تقسیم کردند. آنها خوشه‌های بدست آمده از SOM را با دو الگوریتم سستی وارد و میانگین k مقایسه کردند و دریافتند که SOM نسبت به دو روش دیگر توانایی بیشتری برای شناسایی مناطق همگن دارد. (Ley et al., (2011 از دو نوع ورودی SOM برای مقایسه شباهت‌های هیدرولوژیکی ۵۳ حوضه آبخیز در آلمان استفاده کردند. آنها گروه حوضه‌هایی که با رفتار پاسخ هیدرولوژیکی خوشه‌بندی شده بودند را با خوشه‌هایی که براساس خصوصیات حوضه‌های آبخیز خوشه‌بندی شده‌اند مقایسه کردند. نتایج آنها یک همپوشانی ۶۷٪ بین این دو نوع خوشه‌بندی نشان داد که می‌تواند با استفاده از توپولوژی SOM بهبود یابد. (Chavoshi et al., (2012 به منظور مقایسه روش‌های خوشه‌بندی سخت (سلسله مراتبی و k -میانگین) با روش‌های نرم (C -میانگین و کوهونن) با استفاده از داده‌های ۶۱ حوضه واقع در نوار ساحلی دریای خزر پرداختند. آنها برای آماده‌سازی داده‌های ورودی خوشه‌بندی از تحلیل مولفه اصلی استفاده کردند. مقایسه نتایج با استفاده از آزمون همگنی گشتاورهای خطی نشان داد که خوشه‌های به‌دست آمده با روش‌های نرم همگن‌تر هستند. فرسنادیا و همکاران (۱۳۹۱) از تحلیل خوشه‌ای فازی برای تعیین مناطق همگن استان مازندران استفاده کردند و روند منطقه‌ای بارندگی‌ها در سه منطقه‌ی بدست آمده از الگوریتم خوشه‌بندی فازی را مورد مطالعه قرار دادند.

هرچند SOM در خوشه‌بندی حوضه‌های آبخیز به نواحی همگن هیدرولوژیک با موفقیت استفاده شده است، اما تجزیه و تحلیل نقشه خروجی SOM دشوار است. به‌طور کلی این الگوریتم به ندرت خوشه‌های واضحی در خروجی نشان می‌دهد. به‌منظور حل این مشکل پژوهشگران راهکارهای متفاوتی ارائه کردند. (Lampinen and Oja (1992 یک الگوریتم خوشه‌بندی دو مرحله‌ای ارائه کردند که خروجی SOM اول را به‌عنوان ورودی SOM دوم استفاده کردند. آزمایشات آنها روی داده‌های طبیعی و مصنوعی نشان داد که الگوریتم SOM ترکیبی از قابلیت بیشتری نسبت به الگوریتم میانگین k کلاسیک و SOM معمولی برای تفکیک کلاس‌های مختلف داده‌ها برخوردار است.

(2000) Vesanto and Alhoniemi از الگوریتم خوشه‌بندی سلسله مراتبی و میانگین k برای خوشه‌بندی خروجی SOM استفاده کردند. آنها مهمترین مزیت مدل ارائه شده را کاهش نسبتاً زیاد بارگذاری انجام شده بوسیله الگوریتم‌ها، امکان خوشه‌بندی مجموعه داده‌های بزرگ و بررسی چندین استراتژی پردازش متفاوت در زمان محدود اعلام کردند. (Srinivas et al., (2007 از روش ترکیبی خوشه‌بندی SOM و خوشه‌بندی فازی برای شناسایی مناطق همگن هیدرولوژیک در حوضه‌های آبخیز ایندیانا در ایالات متحده آمریکا استفاده کردند و منطقه مورد مطالعه را به ۷ خوشه همگن تقسیم کردند.

در این پژوهش، از یک تکنیک خوشه‌بندی دو مرحله‌ای برای منطقه‌ای کردن حوضه‌های آبخیز استان مازندران استفاده شده است. در مرحله اول، از SOM برای شکل‌دهی نقشه ویژگی دو بعدی استفاده شد. سپس گره‌های خروجی از SOM با استفاده از الگوریتم c -میانگین فازی^۱ (FCM) خوشه‌بندی شد. در مرحله بعد پس از تعدیل همگنی مناطق به‌دست آمده از خوشه‌بندی، تابع توزیع منطقه‌ای برای هر یک از مناطق همگن به‌دست آمد. در نهایت سیلاب منطقه‌ای برای هر یک از مناطق همگن به‌دست آمد.

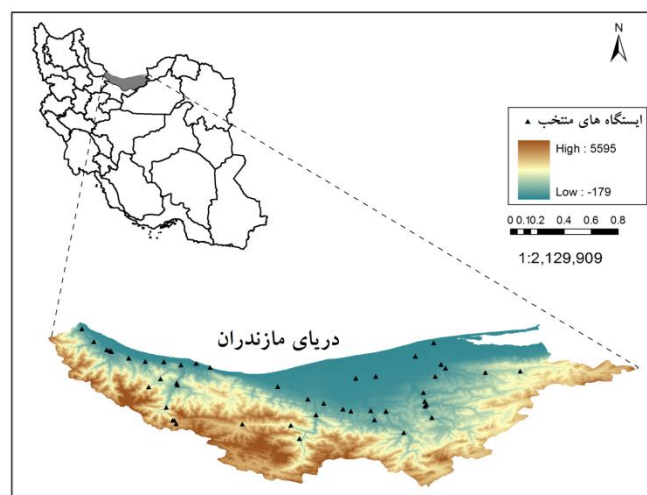
مواد و روش‌ها

منطقه مورد مطالعه و داده‌های مورد استفاده

منطقه مورد مطالعه، استان مازندران در شمال ایران می‌باشد (شکل ۱). این استان با مساحتی بالغ بر ۲۴ هزار کیلومتر مربع در محدوده ۳۵ درجه و ۴۷ دقیقه تا ۳۶ درجه و ۵۸ دقیقه عرض شمالی و ۵۰ درجه و ۳۴ دقیقه تا ۵۴ درجه و ۱۴ دقیقه طول شرقی قرار دارد. ارتفاع استان مازندران از زیر صفر در کناره‌های دریای خزر تا ۵۶۷۰ متر در قله دماوند متغیر است (وزارت نیرو، ۱۳۸۷). میانگین بارندگی در غرب استان مازندران بیشتر از ۱۰۰۰ میلی‌متر و در شرق منطقه بیش از ۳۰۰ میلی‌متر است (وزارت نیرو، ۱۳۸۷). قسمت مرطوب غربی ارتفاع کمی دارد اما بارندگی زیاد در آن باعث مناسب شدن آن منطقه برای فعالیت‌های کشاورزی شده است. برای تحلیل آماری ایستگاه‌های آب‌سنجی منطقه، ابتدا از طریق آب منطقه‌ای استان مازندران آمار حداکثر دبی سالانه و خصوصیات حوضه‌های آبخیز استان

1 - Fuzzy c-mean

مازندران تهیه شد. سپس با توجه به رعایت حداقل طول دوره آماری مورد نیاز، از ۷۸ ایستگاه موجود ایستگاه‌ها با طول دوره آماری کمتر از ۱۲ سال حذف و تعداد ۴۷ ایستگاه برای ادامه مطالعات مورد استفاده قرار گرفت.



شکل ۱- موقعیت استان مازندران و ایستگاه‌های مورد مطالعه

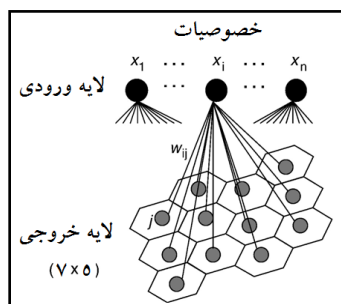
آماده‌سازی داده‌ها

انجام تحلیل مولفه‌ی اصلی (PCA) روی خصوصیات حوضه‌ها قبل از انجام SOM به طرز معنا داری خوشه‌بندی SOM را با کاهش عدم قطعیت در تخمین‌گرهای هیدرولوژیکی در ایستگاه‌های اندازه‌گیری نشده، بهبود می‌بخشد (Di Prinzio et al., 2011). آنها قویا توصیه کردند که قبل از انجام SOM تحلیل مولفه‌ی اصلی روی مجموعه موجود توصیف‌گرهای حوضه انجام شود. لذا در این پژوهش قبل از انجام خوشه‌بندی، تحلیل مولفه اصلی انجام شد.

نگاشت ویژگی خود سامان (SOM)

نگاشت خود سامان تابع چگالی احتمال از داده‌های ورودی تحت یک الگوریتم یادگیری غیر نظارت شونده است، که یک روش موثر در خوشه‌بندی است که برای خلاصه سازی و بصری نمودن داده‌ها به کار می‌رود (Kohonen, 2001). این الگوریتم دارای خصوصیات حفاظت از همسایگی و

تجزیه و تحلیل فضای ورودی متناسب با توزیع داده‌ها را دارا است (Kohonen, 2001, 1982). SOM شامل دو لایه است: یک لایه ورودی شکل گرفته از مجموعه گره‌ها^۱ (یا نرون‌هایی که واحدهای^۲ محاسباتی دارند) و یک لایه خروجی (لایه کوهونن) که توسط گره‌هایی که در یک شبکه دو بعدی قرار گرفته‌اند، تشکیل شده است (شکل ۲).



شکل ۲- طرح شماتیک از ساختار نگاشت ویژگی خود سامان

هر گره در لایه ورودی به هر گره در لایه خروجی توسط شبکه‌ی سیناپسی متصل است. هر گره خروجی دارای یک بردار ضرایب متصل به داده‌های ورودی است. بردار ضرایب یک وزن (یا شدت اتصال) را با نام W بین ورودی و لایه خروجی باز می‌گرداند. وزن‌ها یک شبکه بین واحدهای ورودی (بردار مشخصه‌ها) و واحدهای خروجی وابسته به آنها (گروه‌هایی از بردار مشخصه‌ها) برقرار می‌کند. عملکرد الگوریتم به شرح زیر می‌باشد: زمانی که یک بردار مشخصه ورودی X' به SOM ارائه شد، گره‌ها در لایه خروجی با یکدیگر رقابت می‌کنند و گره برنده (گره‌ای که فاصله‌ی تمامی وزن‌هایش از بردار ورودی حداقل است) انتخاب می‌شود. براساس قاعده یادگیری SOM بردار وزن گره برنده و همسایه‌های از پیش تعریف شده‌اش در الگوریتم، به روز رسانی می‌شوند.

خوشه‌بندی واحدهای SOM

در این پژوهش خروجی SOM با استفاده از ماتریس U و روش خوشه‌بندی فازی، تفکیک شد. ابتدا الگوریتم ماتریس فواصل متحد شده یا ماتریس U (Ultsch, 1993; Ultsch and Siemon, 1990)

1- Nodes
2- Units

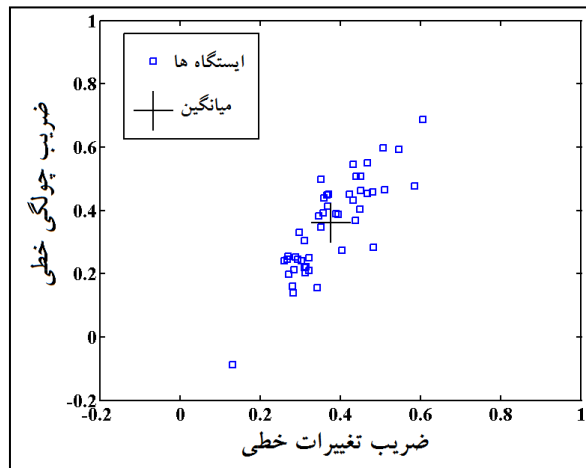
به کار رفت. ماتریس U فواصل بین واحدهای همسایه در نقشه خروجی را محاسبه می‌کند. این فواصل می‌تواند برای آرایه خوشه‌ها با استفاده از مقیاس خاکستری^۱ روی نقشه نشان داده شود (Kohonen, 2001). همچنین ماتریس وزن‌های خروجی SOM به عنوان ورودی در الگوریتم خوشه‌بندی فازی (Bezdek, 1981) برای مشخص کردن خوشه‌ها روی نقشه آموزش یافته به کار رفت. به این منظور شش شاخص تعیین تعداد خوشه بهینه به کار رفت که براساس نتایج دو شاخص ارزیابی خوشه‌بندی فازی زی و بنی تعمیم یافته و وون بهترین نتایج را نشان دادند. همچنین بمنظور بررسی همگنی منطقه‌ای از آزمون‌های مبتنی بر گشتاورهای خطی، شامل آماره ناهمگونی (ناجوری) و آماره ناهمگنی استفاده شد.

نتایج و بحث

تحلیل همگنی ناحیه

ابتدا نمودار گشتاورهای خطی ایستگاه‌های انتخاب شده برای بررسی اولیه منطقه رسم شد (شکل ۳). نمودار ضریب تغییرات خطی نسبت به ضریب چولگی خطی نشان داد که گشتاورهای خطی پراکنده‌اند و استان مازندران نمی‌تواند به عنوان یک منطقه‌ی همگن در نظر گرفته شود. در گام بعدی مقادیر آماره ناهمگنی برای کل استان مازندران به دست آمد. اگر مقدار آماره‌های ناهمگنی کم‌تر از یک باشد، منطقه همگن و اگر بین ۱ تا ۲ باشد، منطقه تا حدی همگن و اگر بزرگ‌تر از ۲ باشد، منطقه کاملاً ناهمگن است. در عمل عنوان می‌شود که معیار $H1$ به منظور بررسی این آزمون مناسب‌تر است (Hosking, 1986). اگر کل استان مازندران را به عنوان یک منطقه‌ی واحد در نظر بگیریم مقدار آماره ناهمگنی برابر $H1 = 2/87$ و $H2 = 1/66$ و $H3 = 0/36$ به دست می‌آید. که نشان‌دهنده ناهمگنی زیاد استان مازندران است و نیازمند تفکیک منطقه به نواحی همگن هیدرولوژیک است.

1- Grey scale



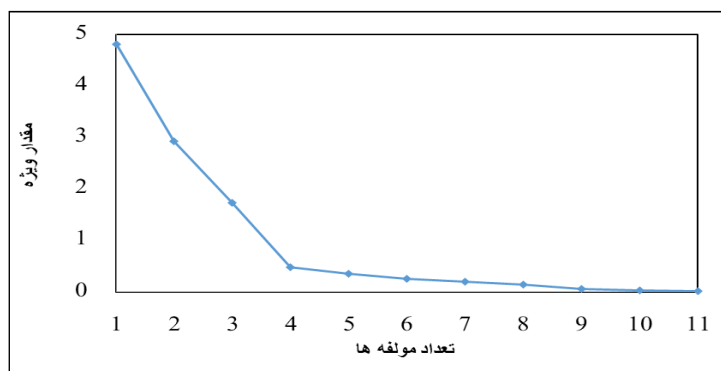
شکل ۳- نمودار نسبت گشتاورهای خطی برای ۴۷ ایستگاه آب‌سنجی در استان مازندران

تعیین مولفه‌های اصلی جهت ورودی الگوریتم SOM

برای کاهش متغیرهای موجود تحلیل مولفه اصلی (PCA) انجام شد. برای مشاهده اینکه PCA یک روش مناسب برای کاهش داده‌ها است، آماره KMO^1 محاسبه شد. آماره KMO برابر 0.713 انجام PCA را روی داده‌های ورودی تایید می‌کند. نمودار سنگ‌ریزه‌ای^۲ (شکل ۴) نشان می‌دهد که تعداد ۴ مولفه بهترین انتخاب است و چهار مولفه اول بیش از ۹۰٪ از کل سهم واریانس مولفه‌های ورودی را نشان می‌دهند. همچنین خصوصیات فیزیوگرافی مساحت حوضه، طول آبراهه، شیب آبراهه اصلی، طول آبراهه اصلی و محیط حوضه که بیشترین تاثیر را در مولفه اول با داشتن ۴۳/۶۷٪ از سهم واریانس متغیرهای ورودی را دارند. به‌علاوه، میانگین ارتفاع حوضه، طول جغرافیایی و میانگین شیب حوضه بیشترین تاثیر در مولفه دوم را با در بر داشتن ۲۶/۵۹٪ از سهم واریانس متغیرهای ورودی دارد. به‌علاوه، مولفه سوم متأثر از میانگین بارندگی سالانه و ارتفاع ایستگاه است. عرض جغرافیایی بیشترین تاثیر را روی مولفه چهارم دارد. بنابر تحلیل فوق ۴ مولفه اول بعنوان ورودی SOM انتخاب شدند.

1- Kaiser-Meyer-Olkin

2- Scree Plot

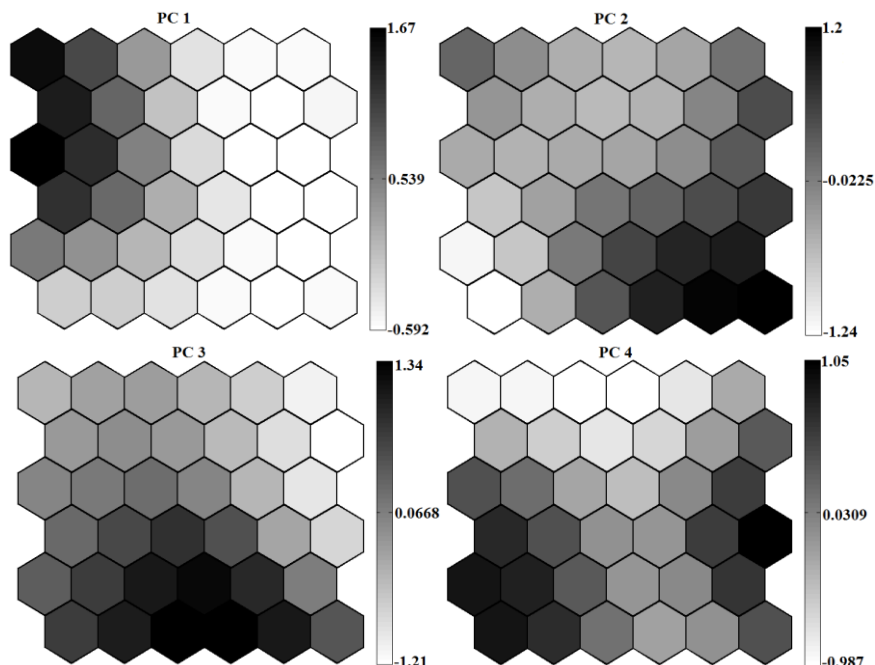


شکل ۴- نمودار سنگریزه‌ای نشان داد که ۴ فاکتور بهترین نتیجه را دارد

خوشه‌بندی SOM

نخست چهار مولفه‌ی اصلی به مدل SOM تغذیه شد. تعداد نرون‌ها در لایه ورودی SOM برابر ۴ (تعداد چهار مولفه‌های اصلی) انتخاب شدند. لایه خروجی براساس پیشنهاد Vesanto et al., (2000) برابر ۳۶ نرون با شبکه‌ی شش وجهی (۶×۶) برای تصویر سازی بهتر تشکیل شد. در شکل (۸) توزیع تدریجی مولفه‌های اصلی در نقشه آموزش دیده SOM نشان داده شده است.

بعد از آموزش SOM الگوریتم ماتریس U بر روی واحدهای آموزش دیده‌ی نقشه SOM انجام شد. ماتریس U هیچ مرز واضحی روی نقشه خروجی SOM نشان نمی‌دهد، لذا بردار وزن‌های SOM برای خوشه‌بندی واحدها با الگوریتم خوشه‌بندی فازی بکار رفت.

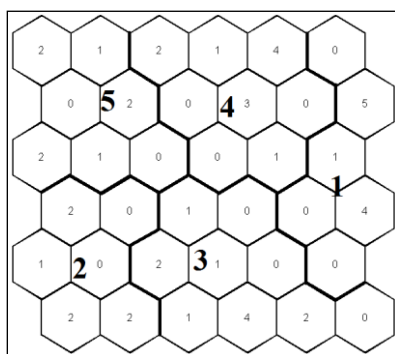


شکل ۵- توزیع تدریجی مولفه‌های اصلی در نقشه آموزش دیده SOM با چهار مولفه‌ی اصلی به عنوان ورودی

خوشه‌بندی فازی واحدهای SOM

بردار وزن‌های SOM به منظور خوشه‌بندی واحدها با الگوریتم c میانگین فازی بکار رفت. تعداد شش شاخص خوشه‌بندی فازی به منظور تعیین تعداد خوشه‌های فازی به کار رفت. شاخص‌های زی و بنی تعمیم یافته و ووان ۵ خوشه‌ی واضح براساس حداقل مقدار شاخص‌ها نشان دادند. شاخص‌های درجه فازی بودن (FPI) و آنتروپی طبقه بندی نرمالیزه (NCE) تعداد خوشه‌ها را به وضوح نشان نمی‌دهند، اما کاهش FPI یکنواختی با افزایش تعداد خوشه‌ها برای $C < 5$ نشان می‌دهد. شاخص‌های آنتروپی افراز PE و ضریب افراز PC، $C = 2$ را به عنوان بهترین افراز نشان دادند. بطور کلی، در $C = 2$ مقدار PE حداقل و PC حداکثر است. این موضوع به این علت است که هر دو شاخص ارتباط مستقیم کمی به جزئیات داده‌ها دارند. در نتیجه، دو شاخص غالباً $C = 2$ را به عنوان افراز بهینه پیشنهاد می‌کنند و این با نتایج Hall and Mins (1999) و Srinivas et al (2008) مطابقت دارد.

براساس نتایج شاخص‌های زی و بنی و ووان $C=5$ انتخاب شد و نقشه‌ی خروجی SOM به پنج خوشه‌ی فازی تقسیم شد (شکل ۶). در شکل (۶) اعداد درون شش ضلعی‌ها نشان دهنده‌ی تعداد ایستگاه‌های اختصاص داده شده به واحدهای SOM بین ۱ تا ۵ است. بعد از انجام خوشه‌بندی آماره همگنی برای هر یک از خوشه‌ها بدست آمد. نتایج در جدول (۱) ارائه شده است. چهار خوشه فازی همگن هستند و یکی از خوشه‌ها کاملاً ناهمگن است.



شکل ۶- خوشه بندی ۴۷ ایستگاه بر روی نقشه آموزش یافته‌ی SOM با استفاده از FCM

زمانی که مجموعه جامعی از متغیرهای تاثیر گذار در تحلیل فراوانی منطقه‌ای سیلاب وجود ندارد، معمولاً مناطق شکل گرفته، برای تحلیل فراوانی منطقه‌ای سیلاب همگن نخواهند بود و نیاز به تعدیل برای بهبود همگنی آن وجود دارد (Hosking and Wallis, 1997). گزینه‌های پیشنهاد شده توسط هاسکینگ و والیس (Hosking and Wallis, 1997) برای اصلاح مناطق ایجاد شده با تحلیل خوشه‌ای عبارتند از:

(۱) حذف یک یا چند ایستگاه از مجموعه داده‌ها؛ (۲) انتقال یک یا چند ایستگاه از یک منطقه به منطقه دیگر؛ (۳) تفکیک منطقه به دو یا چند خوشه جدید؛ (۴) اجازه دادن به یک ایستگاه برای مشارکت در دو یا چند منطقه؛ (۵) انحلال مناطق با انتقال ایستگاه‌هایش به دیگر مناطق؛ (۶) الحاق یک منطقه به منطقه یا مناطق دیگر؛ (۷) الحاق دو یا چند منطقه و تعریف دوباره گروه‌ها؛ (۸) کسب داده‌های بیشتر و تعریف دوباره گروه‌ها.

از بین گزینه‌های مطرح شده فوق، سه گزینه اول در کاستن مقدار شاخص ناهمگنی (H) مفیدتر هستند، حال آنکه ۵ گزینه آخر تضمین کننده حداقل اندازه هر خوشه است. در این پژوهش، گزینه تعدیل کننده (۱) و (۲) برای دستیابی به خوشه‌های همگن به کار رفت. در ابتدا ایستگاه‌هایی که توسط آماره غیر یکنواختی (D) هاسکینگ و والیس غیر یکنواختی شناخته شدند، حذف گردیدند. اگر چه (Hosking and Wallis, 1997, 1993) مقادیر بحرانی آماره غیر یکنواختی ($D > 3$) را برای شناسایی ایستگاه‌های غیرهمگن معرفی کردند، اما بهتر است تمام ایستگاه‌هایی که مقادیر بالای آماره غیریکنواختی را شناسایی کرد. ثانیاً، آماره ناهمگنی منطقه‌ای (H) با حذف و یا تعویض ایستگاه‌ها تعدیل شد. سوماً، ایستگاه‌های غیر یکنواخت، که به میزان معنی داری از کم شدن آماره ناهمگنی منطقه ای (H) ممانعت می‌کنند، شناسایی شده و بعد از حصول اطمینان از زیاد نبودن عضویت فازی این ایستگاه‌ها، حذف شدند.

در جدول (۱) مقادیر آماره ناهمگنی را قبل و بعد از تعدیل مناطق نشان می‌دهد، نتایج نشان می‌دهد که با حذف دو ایستگاه، چهار منطقه‌ی باقی مانده پس از تعدیل به طور قابل قبولی همگن هستند ($H_i > 1$).

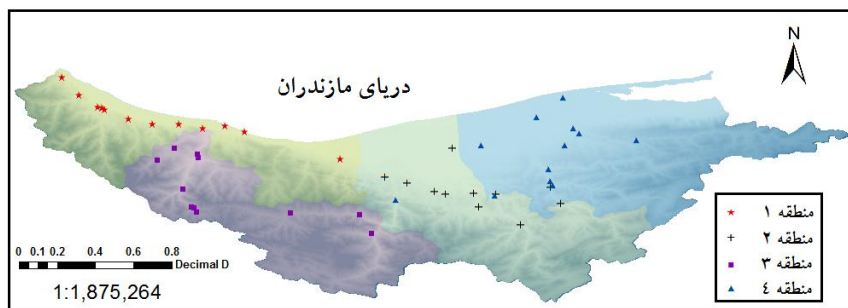
جدول ۱- مقادیر آماره ناهمگنی

قبل از تعدیل		بعد از تعدیل		شماره منطقه
H_i	تعداد ایستگاه	H_i	تعداد ایستگاه	
۰/۹۷	۱۱	۰/۸۷	۱۱	۱
۰/۸۹	۱۱	۰/۸۲	۱۰	۲
۰/۹۸	۱۱	۰/۹۸	۱۱	۳
-	-	۴/۰۱**	۵	۴
۱	۱۲	۰/۹۷	۱۰	۵

** ناهمگن

انتخاب تابع توزیع مناسب برای منطقه همگن

شکل (۷) مناطق همگن نهایی را پس از تعدیل نشان می‌دهد. به منظور انتخاب بهترین توزیع منطقه‌ای، آماره Z^{DIST} برای ایستگاه‌های هر منطقه برای توزیع‌های ۳ پارامتری محاسبه شد (جدول ۲). در این تحقیق پنج تابع توزیع معروف سه پارامتری لجستیک تعمیم یافته، مقادیر حدی تعمیم یافته، لگ نرمال سه پارامتری یا لوگ نرمال سه پارامتری، پیرسون نوع سوم و پارتو تعمیم یافته جهت انتخاب بهترین تابع توزیع منطقه‌ای مورد مقایسه قرار گرفتند. هر چند مقادیر Z^{DIST} در تمامی مناطق، برای چند توزیع آماری کمتر از ۱/۶۴ است، اما توزیع لجستیک تعمیم یافته کمترین مقدار این آماره در هر چهار منطقه داراست.



شکل ۷- مناطق بدست آمده از الگوریتم خوشه بندی SOM و FCM بعد از تعدیل نهایی

جدول ۲- مقادیر آزمون نکویی برازش

Z^{DIST} مقادیر آماره				توزیع آماری
منطقه ۴ (شرق)	منطقه ۳ (ارتفاعات غرب)	منطقه ۲ (وسط)	منطقه ۱ (غرب)	
-۰/۹۸*	۰/۱۲*	-۰/۴۴*	-۰/۵۱*	لوجستیک تعمیم یافته
-۱/۵*	-۰/۵۷*	-۱/۱۳*	-۰/۹۴*	مقادیر حدی تعمیم یافته
-۲/۳۴	-۱/۳۱*	-۱/۷۷	-۱/۷۵	لگ نرمال سه پارامتری
-۳/۷۸	-۲/۵۷	-۲/۸۷	-۳/۱۲	پیرسون نوع ۳
-۳/۱۹	-۲/۵۸	-۳/۰۷	-۲/۴۳	پارتو تعمیم یافته

*توزیع های آماری منتخب

بر این اساس توزیع لجستیک تعمیم یافته به عنوان بهترین توزیع با پارامترهای متفاوت برای هر چهار منطقه شناخته شد و پارامترهای توزیع برای هر منطقه به صورت مجزا با استفاده از روش گشتاورهای وزنی احتمال برآورد شد و در جدول (۳) بیان شده است.

جدول ۳- پارامترهای توزیع لجستیک تعمیم یافته

ξ	α	k	
۰/۷۴۷	۰/۳۱۸	-۰/۳۹۷	منطقه ۱
۰/۸۳۱	۰/۲۸۸	-۰/۳۱۷	منطقه ۲
۰/۸۲۷	۰/۲۷۷	-۰/۳۳۲	منطقه ۳
۰/۷۷۶	۰/۲۹۷	-۰/۳۸۲	منطقه ۴

ξ : پارامتر مکان؛ α : پارامتر مقیاس؛ k: پارامتر شکل.

محاسبه چندک‌های سیلاب با استفاده از روابط منطقه‌ای

در نهایت پس از برآورد پارامترهای تابع توزیع منتخب برای مناطق همگن، دبی سیلاب با دوره بازگشت‌های مختلف برای ایستگاه‌های موجود در خوشه‌های تعدیل شده بوسیله توزیع لجستیک تعمیم یافته برآورد گردید.

نتیجه‌گیری

نتایج نشان داد که الگوریتم خوشه‌بندی ارائه شده قابلیت زیادی در تفکیک حوضه‌های آبریز برای دست‌یابی به مناطق همگن هیدرولوژیکی دارد که در تحلیل فراوانی منطقه‌ای سیلاب تاثیر گذار است. در ذیل خلاصه‌ای از نتایج این تحقیق ارائه شده است.

۱. SOM یک روش مفید برای دستیابی به مناطق همگن هیدرولوژیک است، زیرا یک اجرای عالی از بصری سازی و خلاصه سازی خصوصیات حوضه آبریز و توزیع مولفه‌های اصلی را نشان داد.

۲. هر چند ماتریس U یک روش بصری سازی مناسب برای بازدید ابتدایی از تعداد خوشه‌ها در نقشه‌ی آموزش یافته‌ی SOM است اما یک روش قوی برای مشخص کردن مرز خوشه‌ها روی

نقشه‌ی SOM تلقی نمی‌شود و نیاز به روش‌های دیگر مانند FCM برای تعیین مرز نهایی خوشه‌ها بر روی نقشه‌ی خروجی SOM دارد.

۳. شاخص‌های ارزیابی خوشه‌بندی فازی مانند ضریب افراز و آنتروپی افراز، که با داده‌ها ارتباط مستقیم کمی دارند، در تفکیک مناطق همگن هیدرولوژیکی کارایی ندارند. صرف نظر از مقدار فازی سازی (m) انتخاب شده، اغلب هر دو شاخص تعداد خوشه را برابر ۲ پیشنهاد می‌کنند.
۴. شاخص‌های ارزیابی خوشه‌بندی فازی که همزمان میزان فازی سازی و ساختار داده‌ها را محاسبه می‌کنند (مانند شاخص‌های زی و بنی و وون)، کاملاً در تعیین بهترین افراز با الگوریتم خوشه‌بندی فازی موثراند. این موضوع نتایج (Srinivas et al, (2008) رو تایید می‌کند.

تشکر و قدردانی

پژوهشگران این طرح از معاونت فنی و پژوهش‌های شرکت سهامی مدیریت منابع آب ایران و به‌ویژه شرکت آب منطقه‌ای مازندران برای حمایت مالی، کمال تشکر و قدردانی را دارند.

منابع

- فرسادنیا ف.، رستمی کامرود م.، مقدم‌نیاع. (۱۳۹۱). تحلیل روند بارندگی در استان مازندران با استفاده از روش من-کندال منطقه‌ای. تحقیقات منابع آب ایران، سال ۸، شماره ۲: ۶۰-۷۰.
- مطالعات جامع مهندسی رودخانه‌های استان مازندران، وزارت نیرو، شرکت سهامی آب منطقه‌ای استان مازندران، (۱۳۸۷).
- Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York.
- Burn, D.H. and Goel, N.K. (2000). The formation of groups for regional flood frequency analysis. Hydrological Sciences Journal. 45(1), pp. 97-112.
- Chavoshi, S., Azmin Sulaiman, W.N., Saghafian, B., Sulaiman, MD. N.B., Latifah, A.M. (2012). Soft and hard clustering methods for delineation of hydrological homogeneous regions in the southern strip of the Caspian Sea Watershed. Journal of Flood Risk Management, Volume 5, Issue. 4, pp. 282-294.
- Hall, M.J. and Minns, A.W. (1999). The classification of hydrologically homogeneous regions. Hydrological Sciences Journal. 44 (5), pp. 693-704.

- Hosking, J.R.M. and Wallis, J.R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*. 29, pp. 271–281.
- Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach Based on LMoments*. Cambridge University Press: Cambridge.
- Hosking, J.R.M. (1986). The theory of probability weighted moments. Res. Rep. RC 12210, IBM Research Division, Yorktown Heights, NY.
- Jingyi, Z. and Hall, M.J. (2004). Regional flood frequency analysis for the Gan-Ming River basin in China. *Journal of Hydrology*. 296, pp. 98–117.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*. 43, pp. 59–69.
- Kohonen, T. (2001). *Self-Organizing Maps*. Springer, Berlin, Germany.
- Lampinen, J. and Oja, E. (1992). Clustering properties of hierarchical self-organizing maps. *Journal of Mathematical Imaging and Vision*. 2 (2–3), pp. 261–272.
- Ley, R., Casper, M.C., Hellebrand, H., Merz, R. (2011). Catchment classification by runoff behaviour with self-organizing maps (SOM). *Hydrology and Earth System Sciences*, 15(9), pp. 2947-2962.
- Lin, G. and Chen, L. (2006). Identification of homogenous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*. 324, pp. 1-9.
- Lin, G. and Wang, C. (2006). Performing cluster analysis and discrimination analysis of hydrological factors in one step. *Advances in Water Resources*. 29, pp. 1573-1585.
- Rao, A.R. and Srinivas, V.V. (2006). Regionalization of watersheds by fuzzy cluster analysis. *Journal of hydrology*. 318, pp. 57-79.
- Srinivas, V.V., Tripathi S., Rao A.R., Govindaraju R.S. (2007). Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *Journal of Hydrology*. 348, pp. 148– 166.
- Ulsch, A. and Siemon, H.P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In: *Proceedings of INNC'90, International Neural Network Conference*. Kluwer Academic Publishers, Dordrecht, Netherlands. pp. 305_308.
- Ulsch, A. (1993). Self-organizing neural networks for visualization and classification. In: *Opitz, O., Lausen, B., Klar, R. (Eds.), Information and Classification*. Springer, Berlin. pp. 307_313.

- Vesanto, J. and Alhoniemi, R. (2000). Clustering of the self organizing map. IEEE Trans. Neural. Netw. 11 (3), pp. 586-600.
- Di Prinzio, M., Castellarin, A., Toth, E. 2011. Data-driven catchment classification: application to the pub problem. Hydrology and Earth System Sciences, 15(6): 1921-1935.
- Srinivas, V.V., Tripathi, S., Ramachandra Rao, A. and Govindaraju, R. S. 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. Journal of Hydrology, No. 348, PP. 148-166.

Regionalization of Mazandaran Province's Basin to Flood Forecasting

Abstract

One of the estimation methods of flood quantiles in ungauged watersheds or watersheds with short records is the use of regional frequency analysis method. In studies on regional frequency analysis, clustering methods are used to achieve hydrological homogeneous regions. The self-organization feature map (SOM) has been used recently in several researches for clustering the watersheds. But the key issue in using the SOMs is interpretation of the SOMs output units to find the hydrological homogeneous regions. For this reason, SOM is used as input to other clustering algorithms. In this study, firstly, the SOM was used to generate a two-dimension feature map, and then output nodes of SOM were used by fuzzy c-mean clustering to form the required regions for regional flood frequency analysis. The optimum number of fuzzy clusters was determined six indices of fuzzy clustering evaluation. The results showed that Xie-Beni and Kwon indices have better performance than the other indices. Next, the homogeneity of each of the regions obtained by the combined clustering algorithm was evaluated by the heterogeneity statistic of Hosking and Wallis. ZDIST statistics for the stations of each of the four hydrological homogenous regions were also calculated for the three-parameter distributions and the generalized logistic distribution was selected as the best regional distribution. The parameters of this distribution were calculated by linear moments method, then the floods with different return periods at each of the stations were estimated by using regional equations. According to the results of this study, the proposed algorithm can be used to achieve hydrological homogeneous regions in the other parts of Iran. In addition, a graphical and User-friendly software for SOM and fuzzy clustering and fuzzy clustering indices calculation were developed by the researchers in the Windows operating system.

Keywords: Regionalization, fuzzy c-means clustering, self-organization feature maps, Flood index, linear moments, hydrologic homogeneity